

A Local Least Squares Framework for Ensemble Filtering

Jeffrey L. Anderson
NOAA / GFDL

correspondence to
NCAR/MMM
P.O. Box 3000
Boulder, CO 80307-3000

Submitted to Monthly Weather Review
18 October, 2001

Abstract

A number of methods for using ensemble integrations of prediction models as an integral part of data assimilation have appeared in the atmospheric and oceanic literature. In general, these methods have been derived from the Kalman filter and have been known as ensemble Kalman filters. A more general class of methods that includes these ensemble Kalman filter methods is derived starting from the nonlinear filtering problem. When working in a joint state / observation space, many features of ensemble filtering algorithms become easier to derive and compare. The ensemble filter methods derived here make a (local) least squares assumption about the relation between the prior distributions of an observation variable and the model state variables. In this context, the update procedure applied when a new observation becomes available can be described in two parts. First, an update increment is computed for each prior ensemble estimate of the observation variable by applying a scalar ensemble filter. Second, a linear regression of the prior ensemble sample of each state variable on the observation variable is performed. This regression computes update increments for each state variable ensemble member from the corresponding observation variable increments. The regression can be applied globally or locally using Gaussian kernel methods.

Several previously documented ensemble Kalman filter methods, the perturbed observation ensemble Kalman filter and ensemble adjustment (or ensemble square root) Kalman filter, are developed in this context. Some new ensemble filters that extend beyond the Kalman filter context are also discussed. The two part method provides a computationally efficient implementation of ensemble filters and allows for more straightforward comparison of these methods since they differ only in the solution of a scalar filtering problem.

1. Introduction

Interest in data assimilation methods using ensemble integrations of prediction models is growing rapidly in the atmospheric and oceanic communities. This is occurring because methods for doing assimilation with ensembles are rapidly maturing, and because both prediction centers and research groups are becoming increasingly interested in characterizing more information about the probability distribution of the climate system than can be revealed by a single assimilated state estimate.

Ensemble assimilation methods were originally developed as generalizations of the Kalman filter (Kalman and Bucy 1961, Courtier et al 1993). This led to a sequence of related methods known as ensemble Kalman filters (Evensen 1994) which have been extended to increasingly general assimilation problems (Houtekamer and Mitchell 1998). More recently, other variants, still referred to as ensemble Kalman filters (Bishop et al 2001, Anderson 2001), have appeared in the literature demonstrating improved assimilation error characteristics and/or decreased computational cost. Some of these new methods were developed directly from the probabilistic statement of the nonlinear filtering problem, rather than by generalizing from the Kalman filter. Developing filters in this context can lead to a more straightforward understanding of their capabilities for those not intimately related with the intricacies of the Kalman filter.

Here, a framework is developed in which most ensemble Kalman filter methodologies documented to date can be described while still supporting a more general class of ensemble filters. The derivation begins with the nonlinear filtering problem and applies a sequence of simpli-

fyng assumptions. The introduction of a joint state / observation space (Tarantola 1987) leads to an ability to deal with observations related to the model state variables by non-linear functions. A least squares assumption (equivalent to assuming a local Gaussian relation amongst the prior joint state variables) has been made, sometimes indirectly, in many descriptions of ensemble Kalman filters. Here, that assumption is made explicitly and a significant simplification in the description of the algorithms results. Under the assumptions made here, the ensemble filter problem simplifies to an application of a nonlinear filter to a scalar, followed by a sequence of linear regressions. This simplification makes it easier to analyze the relative capabilities of a variety of ensemble filter implementations and can lead to reduced computational cost.

Section 2 derives this context for ensemble filtering and section 3 shows how several previously documented ensemble Kalman filters are related. Section 4 discusses details of methods for doing scalar assimilation problems. Section 5 briefly discusses the need for limiting the impact of observations on state variables while section 6 offers conclusions.

2. Ensemble Filtering

To simplify notation, this section only discusses what happens at a single time at which observations become available. Discussion of how filter assimilations are advanced in time using ensemble methods and prediction models can be found in Anderson (2001) (hereafter A01), Houtekamer and Mitchell (1998), and Jazwinski (1970). Basically, each ensemble member is integrated forward in time independently using a forecast model between times at which observations are available.

A. Joint state/observation space and Bayesian framework

The joint state/observation space (Tarantola 1987, A01) is defined by the joint space state vector

$$\mathbf{z} = [\mathbf{x}, \mathbf{h}(\mathbf{x})] = [\mathbf{x}, \mathbf{y}] \quad (1)$$

where \mathbf{x} is the model state vector, $\mathbf{y} = \mathbf{h}(\mathbf{x})$, referred to as observation variables, is the expected value of the observations available at this time, and \mathbf{z} is a vector of length $n + m$ where n is the number of state variables and m is the number of observations available at this time. Observations are assumed to be selected from a distribution with expected value given by $\mathbf{h}(\mathbf{x})$ and an associated observational error distribution.

Using Bayesian statistics as in Jazwinski(1970) and A01, the distribution of the posterior (or updated) distribution $\mathbf{z}^u = [\mathbf{x}^u, \mathbf{y}^u]$ can be computed from the prior distribution, $\mathbf{z}^p = [\mathbf{x}^p, \mathbf{y}^p]$, as

$$\mathbf{p}(\mathbf{z}^u) = \mathbf{p}(\mathbf{y}^o | \mathbf{z}^p) \mathbf{p}(\mathbf{z}^p) / \text{normalization} \quad (2)$$

where \mathbf{y}^o is an m -vector of the observed values available at this time. In the ensemble methods applied here, the normalization factor in the denominator is not used explicitly (Anderson and Anderson 1999).

One implication of (2) is that subsets of observations with independent (observational error) distributions can be assimilated sequentially. Let \mathbf{y}^o be composed of s subsets of observations, $\mathbf{y}^o = \{\mathbf{y}_1^o, \mathbf{y}_2^o, \dots, \mathbf{y}_s^o\}$, where the distribution of the observational errors for observations in subset i is independent of the distribution for the observations in subset j , for $i \neq j$. Then

$$p(\mathbf{y}^o | \mathbf{z}^p) = \prod_{i=1}^s p(\mathbf{y}_i^o | \mathbf{z}^p) \quad . \quad (3)$$

In particular, if the individual scalar observations in \mathbf{y}^o have mutually independent error distributions, they can be assimilated sequentially in any order without changing the result in (2). This allows sequential assimilation with observational error distributions represented as Gaussians as long as the observational error covariance matrix is diagonal. This was pointed out by Houtekamer and Mitchell (2001) in the ensemble Kalman filter context and used in A01. Eq. (3) depends only on the observing system and makes no assumptions about the prior joint state distribution or how it is represented.

B. An ensemble method for the filtering problem

In ensemble methods for solving (2), information about the prior distribution of the state variables, \mathbf{x}^p , is available as a sample of size N produced by N applications of a prediction model. An ensemble sample of the prior observation vector, \mathbf{y}^p , can be created by applying the forward observation operator, \mathbf{h} , to each ensemble sample of \mathbf{x}^p .

Some Monte Carlo (ensemble) methods also have a weight, w , associated with each ensemble member. The possibility of weighted ensembles is not discussed in detail here, but the methods in this section are easily generalized to this case. Attempts to apply the most common types of weighting / resampling Monte Carlo algorithms in high-dimensional spaces have faced significant difficulties.

Observational error distributions of climate system observations are generally only poorly known and are often specified as Gaussian with zero mean (known instrument bias is usually corrected by removing the bias from the observation during a pre-processing step). Some observations have values restricted to certain ranges, for instance precipitation must be positive. Redefining the observation variable as the log of the observation can lead to a Gaussian observational error distribution in this case (Tarantola 1987). Given Gaussian observational error distributions, observations can be decomposed into subsets where observational errors for observations in each subset are correlated but observational errors in different subsets are uncorrelated. In other words, \mathbf{R} , the observational error covariance matrix, is block diagonal with each block being the size of the number of observations in the corresponding observation subset. Error distributions for the different subsets are independent, so the subsets can be assimilated sequentially in (2) in an arbitrary order.

For many commonly assimilated observations, each scalar observation has an error distribution that is independent of all others, allowing each scalar observation to be assimilated sequentially (Houtekamer and Mitchell 2001). If the observational covariance matrix is not strictly diagonal, a singular value decomposition (SVD) can be performed on \mathbf{R} . The prior joint state ensembles can be projected onto the singular vectors and the assimilation can proceed using this new basis, in which \mathbf{R}' , the observational covariance matrix, is diagonal by definition. Upon completion of the assimilation computation, the updated state vectors can be projected back to the original state space. Given the application of this SVD, a mechanism for sequential assimilation of scalar observations implies no loss of generality for observations with arbitrary Gaussian error distributions. In everything that follows, results are presented only for assimilation of a single sca-

lar observation so that $m = 1$, with joint state space size $k = n + m = n + 1$. Allowing arbitrary observational error distributions represented as a sum of Gaussians is a straightforward extension to the methods described below.

C. Two step data assimilation procedure

Following A01, define the joint state space forward operations operator for a single observation as the order $1 \times k$ linear operator $\mathbf{H} = [0, 0, \dots, 0, 1]$. The expected value of the observation can be calculated by applying \mathbf{H} to the joint state vector, \mathbf{z} , which is equivalent to applying the possibly non-linear operator \mathbf{h} to \mathbf{x} . The conversion of the possibly non-linear \mathbf{h} to the linear \mathbf{H} is a primary motivation for applying ensemble filters in the joint state space.

The updated probability for the *marginal* distribution of the observation joint state variable, y , can be formed from eq. (2) with the simple form

$$p_y(y^u) = \mathbf{p}(y^o | y^p) \mathbf{p}_y(y^p) / (norm). \quad (4)$$

where the subscript on the probability densities indicates a marginal probability on the observation variable, y . The one dimensional problem for this marginal distribution can be solved by a variety of methods, some of which are discussed in sections 3 and 4. Note that (4) does not depend on any of the model state variables.

This suggests a partitioning of the assimilation of an observation into two parts. The first determines updated ensemble members for the observation variable y given the observation, y^o . To update the ensemble sample of y^p , an increment, Δy_i , is computed for each ensemble member, $y_i^u = y_i^p + \Delta y_i$, $i = 1, \dots, N$ where N is the ensemble size.

Given increments for the observation variable, the second step computes corresponding increments for each ensemble sample of each state variable, $\Delta x_{i,j}$ (i indexes the ensemble member and $j = 1, \dots, k$ indexes which joint state variable throughout this report). This requires assumptions about the prior relationship between the joint state variables. Although reasonable alternatives exist (Tarantola 1987), the assumption used here is that the prior distribution is Gaussian (or a sum of Gaussians which allows generality). This is equivalent to assuming that a least squares fit (local least squares fit) to the prior ensemble members summarizes the relationship between the joint state variables.

Figure 1 depicts the simplest example in which there is only a single state variable, x . The observation variable, y , is related to x by the operator h which is nonlinear in the figure. Increments for each ensemble sample of y have been computed. The corresponding increments for x are then computed by a global least squares fit (linear regression) so that

$$\Delta x_i = \frac{\sigma_{x,y}}{\sigma_{y,y}} \Delta y_i. \quad (5)$$

The change in the i th ensemble sample of the state variable due to observation variable y is equal to the prior covariance of x with y , $\sigma_{x,y}$ divided by the prior variance of y , $\sigma_{y,y}$ times the change in the i th ensemble sample of the observation variable. This is just a statistical linearization and inversion of the operator h . This linearization can be done globally by computing the global sample covariance and using this for the regression for each ensemble member (Fig. 1).

The linearization can also be done locally (Fig. 2) by computing local estimates of covariance for each ensemble member. This can be done, for instance, by only using a set of nearest

neighbors (in y , in x , or in some combined distance metric) to compute sample covariance. Figure 2 shows an idealized form of nearest neighbor linearization in which only a single closest ensemble member is used to compute the statistical linearization. Related methods for doing local Gaussian kernel approximations of this type can be found in Silverman (1986) and Bengtsson (2001). When x is functionally related to y as in Fig. 2, local linearization methods like this can give significantly enhanced performance when h is strongly non-linear over the prior ensemble range of y .

If h is nonlinear as in the figures, the statistical linearization is only valid locally. To minimize errors due to the linearization, whether global or local linearizations are applied, it is desirable that the observation variable increments, Δy_i , should be as small as possible. This is discussed further in section 4C.

This two step method can be extended trivially to problems with arbitrary numbers of state variables. Least squares fits can be made to compute the increments for each state variable, x_j , $j = 1, \dots, n$, given the increments in the ensemble members for y . When a global linearization is applied using a least squares fit, a single Gaussian is assumed to approximate the prior relation of the variables. The increments, $\Delta x_{i,j}$, for each ensemble sample of each state variable in terms of Δy_i can be computed independently by regression:

$$\Delta x_{i,j} = \frac{\sigma_{x_j, y}}{\sigma_{y, y}} \Delta y_i. \quad (6)$$

Again, local linearizations could be performed using Gaussian (or extended Gaussian) kernel methods (Tarantola 1987) in which only some subset of local information is used to compute the covariance from the ensemble sample. In (6), all relevant information about the prior covariance of the model state variables, \mathbf{x} , needed to compute increments is contained in the correlation of the individual scalar state variables with the observation variable y .

When the state variable being updated and the observation variable are not functionally related as is possible in the case of multivariate model state, the use of local linearizations can be more problematic. Figure 3 shows an example where state variable x_1 is being updated by an observation, y^o . The expected value of the observation is $y = h(x_2)$, where x_2 is a second state variable, here moderately correlated with x_1 . In this case, the linear regression for x_1 performs a statistical linearization in the presence of noise. Using large (global) regressions is useful to filter out this noise. On the other hand, using local linearizations can help to resolve more of the structure of h . Applying local regressions that are based on too few ensemble members can lead to disastrous over-fitting behavior as demonstrated by the application of an idealized single nearest neighbor linearization in Figure 3. Appropriate trade-offs in choosing local versus global linearizations are an important part of tuning ensemble filters for improved performance.

3. Relation to Ensemble Kalman Filters

A variety of ensemble Kalman filters have been described in the literature (see for example Evensen and van Leeuwen 1996, Keppenne 2000, Mitchell and Houtekamer 2000). Other closely related methods for doing assimilation have been described by Lermursaux and Robinson (1999) and Miller et al (1994). This section demonstrates that the two most common of these, the perturbed observations ensemble Kalman filter (EnKF) and the Ensemble Adjustment Kalman Filter (EAKF) can be recast in the two step framework outlined in the previous section. At the heart of these ensemble Kalman filters is the fact that the product of the joint prior Gaussian with

mean $\bar{\mathbf{z}}^p$, covariance Σ^p and weight w and the Gaussian observation distribution with mean \mathbf{y}^o and error variance \mathbf{R} has covariance

$$\Sigma^u = [(\Sigma^p)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1}, \quad (7)$$

mean

$$\bar{\mathbf{z}}^u = \Sigma^u [(\Sigma^p)^{-1} \bar{\mathbf{z}}^p + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o], \quad (8)$$

and an associated relative weight

$$\mathbf{D} = \exp\{ -[(\bar{\mathbf{z}}^p)^T (\Sigma^p)^{-1} \bar{\mathbf{z}}^p + (\mathbf{y}^o)^T \mathbf{R}^{-1} \mathbf{y}^o - (\bar{\mathbf{z}}^u)^T (\Sigma^u)^{-1} \bar{\mathbf{z}}^u] / 2 \} w, \quad (9)$$

as in A01.

The discussion that follows assumes sequential assimilation of scalar observations so that in (7-9), \mathbf{y}^o is a vector of length 1 and \mathbf{R} is a 1x1 matrix. Additional simplifications in computing the product of Gaussians can then be made easily. The order of the prior joint state covariance, Σ^p , is $k \times k$ where $k = n+1$. The updated covariance from (7) can be written

$$\Sigma^u = [\mathbf{I} - (r + \sigma_{k,k})^{-1} \Sigma_{0k}^p] \Sigma^p \quad (10)$$

where $\sigma_{k,k}$ is the prior observation variable error variance (the k th diagonal element of Σ^p), r is the observation error variance (the only element of the 1x1 matrix \mathbf{R}) and Σ_{0k}^p is the matrix consisting of Σ^p with all elements except those in the last column set to 0. The last column is the prior covariance of each joint state variable with the observation variable. The change in the covariance due to the assimilation of a single observation is then

$$\Delta \Sigma = \Sigma^u - \Sigma^p = -(r + \sigma_{k,k})^{-1} \Sigma_{0k}^p \Sigma^p. \quad (11)$$

Substituting (10) into the expression for the updated mean from (8) gives

$$\bar{\mathbf{z}}^u = [\mathbf{I} - (r + \sigma_{k,k})^{-1} \Sigma_{0k}^p] \Sigma^p [(\Sigma^p)^{-1} \bar{\mathbf{z}}^p + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o]. \quad (12)$$

Noting that $\Sigma^p \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o = \Sigma_{0k}^p \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o$ and $\Sigma_{0k}^p \Sigma_{0k}^p = \sigma_{k,k} \Sigma_{0k}^p$, this becomes

$$\bar{\mathbf{z}}^u = [\mathbf{I} - (r + \sigma_{k,k})^{-1} \Sigma_{0k}^p] \bar{\mathbf{z}}^p + r^{-1} [\mathbf{I} - (r + \sigma_{k,k})^{-1} \sigma_{k,k}] \Sigma_{0k}^p \mathbf{H}^T \mathbf{y}^o. \quad (13)$$

An equation for the change in the mean due to assimilating the observation is

$$\Delta \bar{\mathbf{z}} = \bar{\mathbf{z}}^u - \bar{\mathbf{z}}^p = (r + \sigma_{k,k})^{-1} \Sigma_{0k}^p (\mathbf{H}^T \mathbf{y}^o - \bar{\mathbf{z}}^p) = (r + \sigma_{k,k})^{-1} (y^o - \bar{z}_k^p) [\sigma_{1,k}, \sigma_{2,k}, \dots, \sigma_{k,k}]^T \quad (14)$$

where the single element of the vector \mathbf{y}^o is y^o and \bar{z}_k^p is the prior mean value of the observation variable.

Finally, in this case the weight \mathbf{D} from (9) (a scalar) depends only on the observation and the observation variable and simplifies to

$$\mathbf{D} = \exp \left\{ -(y^o - \bar{z}_k^p)^2 / 2(r + \sigma_{k,k}) \right\} w. \quad (15)$$

It is easily verified that computing the impact of the observation on each state variable independently in (11), (14) and (15) is equivalent to computing the impact on all state variables at once. This was pointed out, but not rigorously derived in A01 where assimilations were performed by looking at the impact of an observation on each state variable in turn.

A. Perturbed Observations Ensemble Kalman Filter

In its traditional implementation, the perturbed observation ensemble Kalman filter (Houtekamer and Mitchell 1998) (EnKF) uses a random number generator to sample the observational error distribution (specified as part of the observing system) and adds these samples to the observation, \mathbf{y}^o , to form an ensemble sample of the observation distribution, \mathbf{y}_i^o , $i = 1, \dots, N$. In most implementations (Houtekamer and Mitchell 1998), the mean of the perturbations is adjusted to be 0 so that the perturbed observations have mean \mathbf{y}^o ; other clever methods for perturbing the observations can preserve other aspects of the distribution (Pham 2001) (the discussion below applies whether adjustment to the means or other types of perturbation algorithms are applied or not). Σ^p is computed using sample statistics from the prior joint state ensemble and (7) is computed once to find the value of Σ^u . Eq. (8) is then applied N times with $\bar{\mathbf{z}}^p$ replaced with \mathbf{z}_i^p and \mathbf{y}^o replaced by \mathbf{y}_i^o in the i th application to compute N ensemble members for \mathbf{z}^u . This method is described using more traditional Kalman filter terminology in Houtekamer and Mitchell (1998). As shown in Burgers et al (1998), computing a random sample of the product as the product of random samples is a valid Monte Carlo approximation to the nonlinear filtering equation (2). Note that all ensemble members are assumed equally weighted in both the EnKF and EAKF so that (9) is not used, however, (9) may be relevant for other ensemble filtering methods (see section 4).

An equivalent two step procedure for the EnKF begins by computing the update increments for the observation variable, y , which can be done as a scalar problem independent of the other joint state variables as indicated by (4). Perturbed observations are generated as above and (7) is used to compute an updated variance for the observation variable; all matrices here are order 1×1 so inverses are simply reciprocals. Eq. (8) is evaluated N times to compute y_i^u , with $\bar{\mathbf{z}}^p$ and \mathbf{y}^o replaced by y_i^p and y_i^o , where the subscript refers to the value of the i th ensemble member.

Equation (14) applies in this case, since the updated covariance in the full dimension EnKF is computed by (7), and can be used to compute the increments for all other state variables given the value of $\Delta y_i = y_i^u - y_i^p$ (which is by definition the k th component of the k -vector $\Delta \mathbf{z}_i$). All components of $\Delta \mathbf{z}_i$ can be computed from (14) as

$$\Delta z_{i,j} = \frac{\sigma_{j,k}}{\sigma_{k,k}} \Delta y_i. \quad (16)$$

where the first subscript on $\Delta \mathbf{z}$ indexes which ensemble member and the second indexes which state variable. This is the regression formula (6) presented in section 2C derived from the assumption of a Gaussian relation between the prior state variables. Implementing the EnKF in this two step fashion gives results identical (to computational round-off) to previous implementations of the EnKF but at reduced computational cost.

B. Ensemble Adjustment Kalman Filter / Square Root Filter

A01, Bishop et al (2001), and Hamill and Whitaker (2001) have all described another ensemble filtering method, referred to as the Ensemble Adjustment Kalman Filter (EAKF) by A01. Tippett (personal communication, 2001) providing an analysis of the work of these different authors, points out that the methods are roughly equivalent, and suggests that the name square root filter (Andrews 1968) may be more appropriate.

The EAKF constructs an updated ensemble with a mean and sample variance that exactly satisfy (7) and (8). In A01 this is done by shifting the mean of the ensemble and then adjusting the spread of the ensemble around the updated mean using a linear operator,

$$\mathbf{z}_i^u = \mathbf{A}(\mathbf{z}_i^p - \bar{\mathbf{z}}^p) + \bar{\mathbf{z}}^u \quad (17)$$

where \mathbf{A} is a linear operator that satisfies $\Sigma^u = \mathbf{A}\Sigma^p\mathbf{A}^T$.

The EAKF can be recast in the two step framework developed in section 2. Again, (4) implies that the observation variable y can be updated independently of the other joint state variables. Recalling that y is the k th element of the k -dimensional joint state vector, the updated variance for y can be written

$$\sigma_{k,k}^u = [(\sigma_{k,k}^p)^{-1} + r^{-1}]^{-1} \quad (18)$$

using a scalar application of (7). Applying a scalar version of (8) to compute the updated mean, \bar{y}^u , the updated value of y can be written

$$y_i^u = \alpha(y_i^p - \bar{y}^p) + \bar{y}^u \quad (19)$$

where

$$\alpha = [\sigma_{k,k}^u(\sigma_{k,k}^p)^{-1}]^{1/2} = [r(r + \sigma_{k,k}^p)^{-1}]^{1/2}.$$

For the change in the mean values, $\Delta\bar{\mathbf{z}} = \bar{\mathbf{z}}^u - \bar{\mathbf{z}}^p$, equation (14) holds for the EAKF and this implies that (16) can be used to compute the changes in the state variables by regression from the change in the mean for y .

It can also be shown that the regression formula can be applied for the adjustment of the ensemble members around the mean in the EAKF. Eq. (10) can be rewritten by taking the square root of the operator as

$$\Sigma^u = \mathbf{B}\Sigma^p\mathbf{B}^T \quad (20)$$

where

$$\mathbf{B} = (\mathbf{I} + \beta\Sigma_{0k}^p) \quad (21)$$

and

$$\beta = (\sigma_{k,k}^p)^{-1} \left[\sqrt{r(r + \sigma_{k,k}^p)^{-1}} - 1 \right]. \quad (22)$$

If the deviation of the observation variable around the mean is updated as in (19), applying the linear regression (16) for state variable j gives

$$\Delta z_{i,j} = \sigma_{j,k}(\sigma_{k,k})^{-1} \Delta y_i = \sigma_{j,k}(\sigma_{k,k})^{-1}(\alpha - 1)y_i^p. \quad (24)$$

Then,

$$z_{i,j}^u = z_{i,j}^p + \sigma_{j,k}^p \Gamma y_i^p \quad (25)$$

where

$$\Gamma = (\sigma_{k,k}^p)^{-1}(\alpha - 1) = (\sigma_{k,k}^p)^{-1} \left[\sqrt{r(r + \sigma_{k,k}^p)^{-1}} - 1 \right]. \quad (26)$$

Writing this in vector form for the updates of all joint state variables gives

$$\mathbf{z}_i^u = \mathbf{A}(\mathbf{z}_i^p - \bar{\mathbf{z}}^p) \quad (27)$$

where

$$\mathbf{A} = \mathbf{I} + \Gamma \Sigma_{0k}^p. \quad (28)$$

The matrix \mathbf{A} used in the regression update is identical to \mathbf{B} in (22) demonstrating that using regression in the two step framework is identical to the implementation of the EAKF described in A01.

In summary, the EnKF and EAKF assume a Gaussian relation between the variables in the joint state space prior distribution. However, these methods do not use (7) and (8) directly to compute an updated distribution. Both methods can be recast in terms of the two step assimilation context developed in section 2C. First, update increments are computed for each ensemble sample of the observation variable using scalar versions of the traditional algorithms. Once increments for ensemble samples of the observation variable have been computed, (16) is used to solve for the increments, $\Delta z_{j,i}$, for each state variable in turn in terms of Δy_i by linear regression.

There are a number of implications about the computational complexity of ensemble (Kalman) filtering that can be drawn from (16). First, there is no need to compute the prior covariance among the model state variables (only the prior cross covariance of each state variable with the observation variable is needed, along with the variance of the observation variable) or the complete updated covariance, Σ^u , (only the updated variance of the observation variable is needed). Second, once the observational variables are updated, the increments for the state variables depend only on ratios of prior (co)variances. Any multiplication of the prior covariance matrix by some type of covariance inflation factor as is done in many existing ensemble Kalman filter implementations (Anderson and Anderson 1999, Whitaker and Hamill 2001) does not impact the solution to (16). The impacts of covariance inflation are still felt in the first step in which the increments of the observational variable are computed. Note that many types of covariance inflation directly increase the spread of the joint prior distribution.

C. Degenerate prior covariance

A potential complication occurs if Σ^p is degenerate so that its inverse is not defined. There are several reasons why Σ^p might be degenerate. First, it is possible that the joint state vector is of order greater than the size of the space that it spans. The most obvious instance occurs when an observation is a linear function of the prior model state variables (for instance if a state variable is

observed directly). Second, details of the sample statistics of the prior ensemble could lead to degeneracy. For instance, if the ensemble is smaller than the size of the joint state space Σ^p computed from the sample statistics must be degenerate. Third, details of the prediction model could lead to states that are confined to some sub-manifold of the model state space.

The two-step procedures for the EnKF and EAKF continue to be equivalent to the traditional implementations even if Σ^p is degenerate. In this case, (7) through (9) must be modified to replace the inverses with pseudo-inverses. Let \mathbf{U}^T be an $\eta \times k$ orthogonal matrix whose rows are the set of left singular vectors of Σ^p corresponding to non-zero singular values (η is the rank of Σ^p). Applying \mathbf{U}^T to a vector in the original space gives the projection of that vector on the range of Σ^p . The projection of the covariance on this space is $\mathbf{U}^T \Sigma^p \mathbf{U}$. The probability density of the prior distribution outside of this subspace is zero, so the updated distribution and hence $\Delta \mathbf{z}$ must lie in this subspace. Note that by definition of the SVD, the projection of the covariance on the range of Σ^p projected back into the original space is $\mathbf{U} \mathbf{U}^T \Sigma^p \mathbf{U} \mathbf{U}^T = \Sigma^p$.

Next, define an $\eta \times \eta$ orthogonal matrix \mathbf{B}^T that performs a change of basis in the reduced SVD subspace. Let the $\eta \times k$ matrix $\mathbf{C}^T = \mathbf{B}^T \mathbf{U}^T$ be defined so that the last row of \mathbf{C}^T is the projection of the observation variable, $\mathbf{z}_k = \mathbf{y}$, on the range of Σ^p ; the last column of \mathbf{C}^T is $[0, 0, \dots, 0, \psi]^T$. \mathbf{C}^T exists as long as the observation vector does not lie in the null space of Σ^p , but this must be the case for (2) to have a relevant solution. If the observation vector did lie entirely in the null space, then $\mathbf{p}(\mathbf{y}^0 | \mathbf{z}^p)$ would be 0 with probability 1 and the result of (2) would be a delta function indicating a deterministic solution (and probably an improperly defined problem, Tarantola (1987)).

Within the subspace with the basis given by the rows of \mathbf{C}^T , the projection of the prior covariance has an inverse and (7) through (9) can be applied. The results can then be projected back to the original space:

$$\Sigma'' = \mathbf{C}[(\mathbf{C}^T \Sigma^p \mathbf{C})^{-1} + \mathbf{C}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{C}]^{-1} \mathbf{C}^T. \quad (29)$$

In this subspace, define $\Sigma'^u = \mathbf{C}^T \Sigma^u \mathbf{C}$, $\Sigma'^p = \mathbf{C}^T \Sigma^p \mathbf{C}$, $\mathbf{z}'_i = \mathbf{C}^T \mathbf{z}_i$ (an η -vector), $y'_i = \psi y_i$ is the last element of \mathbf{z}'_i , and $\sigma'^p_{\eta, \eta} = \psi^2 \sigma^p_{k, k}$ is the prior variance of y'_i . Also note that $\mathbf{C}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{C}$ is a $\eta \times \eta$ matrix with all elements 0 except the last column of the last row which is $r'^{-1} = \psi^2 r^{-1}$. Finally, define \mathbf{R}'^{-1} as the 1×1 matrix with only element r'^{-1} and \mathbf{H}' as the η -vector $[0, 0, \dots, 0, 1]$.

In the subspace, (10) through (16) hold for the primed quantities just defined. In particular, (16) in vector form gives

$$\Delta \mathbf{z}'_i = (\sigma'^p_{\eta, \eta})^{-1} \Sigma'^p [0, 0, \dots, 0, \Delta y'_i]^T. \quad (30)$$

Converting this back to the original subspace gives

$$\Delta \mathbf{z}_i = \mathbf{C} \Delta \mathbf{z}'_i = (\sigma'^p_{\eta, \eta})^{-1} \mathbf{C} \Sigma'^p [0, 0, \dots, 0, \psi \Delta y_i]^T. \quad (31)$$

Using the fact that $\mathbf{C}^T \mathbf{C} = \mathbf{I}$,

$$\Delta \mathbf{z}_i = \psi^{-2} (\sigma^p_{k, k})^{-1} \mathbf{C} \Sigma^p \mathbf{C}^T \mathbf{C} [0, 0, \dots, 0, \psi \Delta y_i]^T = (\sigma^p_{k, k})^{-1} \Sigma^p [0, 0, \dots, 0, \Delta y_i]^T, \quad (32)$$

which is a vector form of (16).

This demonstrates that doing the two step procedure in the original space, even if Σ^P is degenerate, gives results corresponding to those given by the previously documented versions of the EnKF and EAKF.

4. Additional Methods for updating the observational variable ensembles

Section 2 developed a framework in which state variables in ensemble filters can be updated once increments to the observational variables have been computed. Section 3 demonstrated that the EnKF and EAKF could be formulated in this framework. In this section, some additional methods for updating the observational variable are discussed. Independence of observational errors continues to be assumed throughout this section so that observations can be processed sequentially. Once update increments for the observation variable are computed by one of the following methods, the rest of the joint state variables can be updated by linear regression using (16).

A. A kernel filter

If the prior distribution of the observation variable may have significant non-Gaussian structure, a kernel method similar to the one employed in Anderson and Anderson (1999) may be useful for computing the update increments. One simple example of kernel methods is the Fukunaga method (Silverman 1986) applied in one dimension. In this algorithm, the prior distribution is represented as a sum of Gaussians with identical variance but different means. The means are the individual prior ensemble samples and the variance is the prior sample variance multiplied by a scaling factor, η . The prior distribution is then

$$p(y^P) = \sum_{i=1}^N N(y_i^P, \eta \Sigma^P)$$

where $N(a, b)$ is a Gaussian with mean a and variance b .

The product of a prior expressed as a sum of Gaussians and a Gaussian observational distribution is equal to the sum of the products of the individual prior Gaussians and the observational Gaussian. The variance of all Gaussians summed in the product is identical in this case and can be computed by a single scalar application of (7). The means will all be different and can be computed by N scalar applications of (12). In the most naive application of this method, an updated ensemble can be generated from this continuous representation by randomly sampling the sum of Gaussians as in Anderson and Anderson (1999).

This kernel method can be extended in a number of ways by allowing more general kernels. For instance, kernels with different means and different variances can be used following a variety of techniques like the class of nearest neighbor methods (Silverman 1986, Bengtsson 2001). In addition, kernels from the class of ‘generalized’ Gaussians as described in Tarantola (1987) can lead to a variety of related kernel algorithms.

B. Quadrature product methods

Update methods that are based directly on ‘quadrature’ solutions to (4) can also be used to find increments for observation variables. One implementation of such a method could begin by computing a continuous approximation to the prior distribution from the ensemble sample; again,

kernel methods are an example. Quadrature methods can then be used to divide the real line into a set of intervals over which the product in the numerator of (4) is computed to approximate the updated distribution. An appropriate method can then be used to sample this updated distribution to generate new ensemble methods.

C. General requirements for an observation variable update

Several characteristics may be important for algorithms used to update the observation variables. First, low quality observations should have small impacts on the ensemble. For atmospheric and oceanic models, the prior distributions may be sampling model ‘attractors’ that have a great deal of structure. Allowing low impact observations to change the ensembles has potential to destroy valuable information. Pure resampling algorithms would be an example of an undesirable method. In this case, the prior ensemble would be converted to a continuous representation which would then be only subtly modified by a low information observation. This updated continuous distribution would then be resampled to generate an ensemble, leading to possibly large increments to ensemble members. The ensemble kernel filter as described above suffers from this deficiency and generally produces assimilations with larger ensemble mean error than do the EnKF and EAKF. This is despite the fact that in many instances it produces more accurate samples of the updated observation variable distribution when the prior is significantly non-Gaussian. Modifications to the kernel filter that limit the impact of low information observations are required to make this method more generally useful.

For related reasons, it is desirable to limit the size of the increments for observation variables as noted in section 2. Since the regression used to update the state variables is a statistical linearization, it is likely to be an increasingly poor approximation as the increments increase. For instance, the updated mean and covariance of the observation variable for the EnKF would be the same if the pairing between the updated, z_k^u , and prior, z_k^p , observational variable ensemble members was changed before the computation of update increments, Δz_k . In some applications, the performance of the EnKF can be dramatically improved by pairing the updated observational variable ensemble members with the prior members in order to reduce the value of the update increments. The most obvious way to do this is to sort the prior and updated observational variable ensemble members and to associate the n th sorted updated ensemble member with the n th sorted prior ensemble member. Doing this reduces much of the difference between the EnKF and EAKF reported in A01. Doing ordered pairing also significantly improves the performance of kernel filter algorithms (section 4A) by reducing the size of increments.

5. Limiting the impact of observations

Section (2C) has shown an efficient method for allowing an observation to impact individual state variables in an ensemble filter. The appropriate solution might appear to be allowing each observation to impact each state variable. However, when the size of affordable ensembles, N , is small, or when N is small compared to the number of observations or the number of state variables, there are several good reasons to limit the impact of an observation on some state variables.

One reason for limiting the impact of observations is simply to reduce computational cost. By limiting the impact of an individual observation to a subset of state variables with which it is

expected to be closely related, the number of cross covariance terms that need to be computed can be greatly reduced.

Hamill et al (2001) have discussed related problems that occur when spuriously large prior correlation magnitudes between an observation variable and state variable occur due to sampling errors. They point out that if there are a large number of observations that are expected to be weakly correlated with a state variable, the impact of spurious correlations may mask the impact of a smaller set of observations with relatively large expected correlation magnitude. This is particularly problematic if the number of observations is large and the ensemble size is small. Hamill et al develop their discussion in a context where there is a natural definition of distance (physical distance in their case) between observations and state variables and suggest reducing the impact of physically distant observations on a state variable by reducing the sample correlation cross covariance between observations and state variables as the distance between them increases. Their results also suggest that it may be difficult to apply ensemble filters with small sample sizes in systems where a state variable is expected to be weakly related to a large number of observations, but strongly related to no observations. For instance, in the case of models of a fluid on the sphere, if the state variables are some sort of basis functions with global support, and observations are primarily of some quantity at a point, one might expect difficulties with spurious correlations for small ensembles.

A final reason for limiting the impact of observations is applicable when the computationally affordable ensemble size is small compared to the number of state variables. As pointed out in section 3C, in such cases the prior covariance matrix, Σ^P , is singular and the increment in the model state will be confined to the range of Σ^P . If the prior state estimate has errors that project onto the null space of Σ^P , these cannot be corrected. This problem can be addressed by allowing different weighted subsets of observations to impact different state variables. The result is that the size of the subspace of the state space over which possible updates occur can be greatly increased; if each state variable sees a different weighted subset of observations, the space spanned by the possible updates may be the complete state space.

This becomes an even more serious problem in atmospheric and oceanic data assimilation when the assimilating model is imperfect, especially when observations of the physical system are being assimilated for prediction purposes. In such cases, the assimilating model may evolve on a sub-manifold of the state space in a fashion that is not consistent with the physical system from which observations are being taken.

Methods for reducing the impact of observations as a function of distance have been used in a variety of data assimilation methods (Gaspari and Cohn 1999) including ensemble filters (Houtekamer and Mitchell 2001, Hamill et al 2001). These studies have indicated the importance of smoothly reducing the influence of observations as a function of (physical) distance to avoid having large differences in the impact of a set of observations on pairs of state variables that are closely related in the prior distribution. Gaspari and Cohn (1999) develop a number of functions with compact support that can be used to reduce the cross covariance of prior state and observation variables in ensemble filters. The compact support of these functions leads to computational efficiency since observations beyond a certain distance do not impact a given state variable.

Houtekamer and Mitchell (2001) applied distance dependent covariances in a type of ensemble filter. In the context developed here, their use of the Schur product can be replaced by a multiplication of the prior joint covariance by a function of the position of the observation and

state variable. Identical results can be obtained in the two step context developed here at a much reduced cost.

6. Conclusions

A local least squares framework for ensemble filtering has been derived leading to a two-step ensemble filtering update procedure when a new observation becomes available. The first step is to compute update increments for each ensemble member to a prior estimate of the value of the observation. This can be done using a variety of algorithms including the perturbed observation ensemble Kalman filter and the ensemble adjustment Kalman filter. Other update methods, for instance a kernel filter, extend beyond the Kalman filter context and can be referred to more generally as ensemble filters.

The second step is to compute increments for each ensemble member of the prior estimate for each state variable in turn. This is done by using the prior ensemble sample to do a linear regression of each state variable in turn on the observation variable. The increments for a given state variable are computed by multiplying the corresponding observation variable increment by the prior covariance of the state and observation variable and dividing by the prior variance of the observation variable.

Deriving a class of ensemble filters in this two step context has a number of advantages. First, it is computationally more efficient than previous descriptions of ensemble Kalman filter algorithms in the literature. The cost is expected to be dominated by the computation of the prior sample cross covariance of the observation and state variables and the variance of the observation variables. A second advantage is that much more elaborate and expensive ensemble update methods can be applied because they need only be applied in a scalar fashion to the observation variables. A final advantage is that it is easier to understand differences between various filtering algorithms. Differences need only be explored in a scalar context making the relative features of, for instance, the EnKF and EAKF much easier to understand.

By lowering the cost and opening up a variety of new filter update algorithms, it is hoped that this local least squares framework can accelerate the development of ensemble filtering algorithms that are best suited for applications such as numerical weather prediction and ocean state estimation.

Acknowledgements: The author is grateful to Chris Snyder, Jim Hansen, Jeff Whitaker, Tom Hamill, Joe Tribbia and Ron Errico for ongoing discussions of ensemble filtering methods. Special thanks are due to Stephen Anderson whose insight led the author to the methods developed here.

Figure Captions

1. An idealized representation showing the relation between update increments for a state variable, x , and an observation variable, y , for a five member ensemble represented by asterisks. The projection of the ensemble on the x and y axes is represented by '+' and the observation, y^o is represented by 'x'. In this case, y is functionally related to x by h . The grey dashed line shows a global least squares fit to the ensemble members. Update increments for ensemble members 1 and 4 for y are shown along with corresponding increments for the ensemble as a whole (thin vectors parallel to least squares fit) and for the x ensemble.

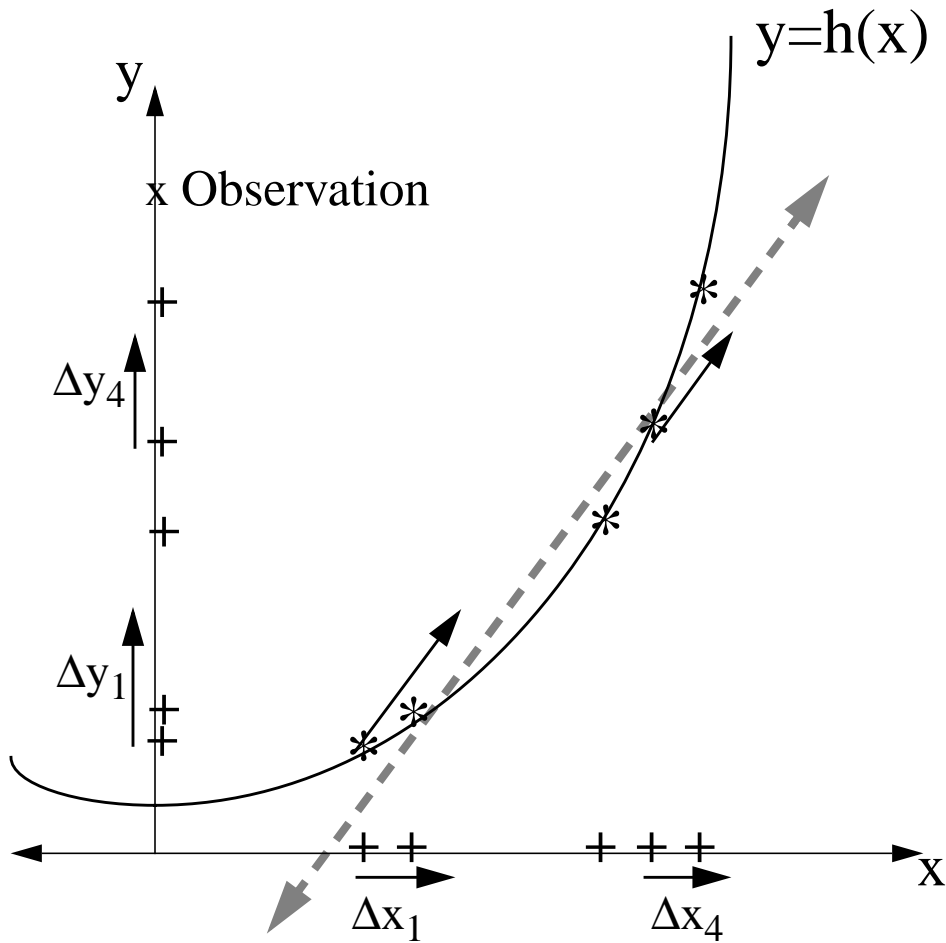
2. As in Figure 1 but showing the application of local least squares fits, in this case using only the nearest neighbor in y , to compute the updates for x given the updates for y . The local updates for the first and fourth ensemble members are shown by the black vectors.

3. As in Figure 1, but now $y = h(x_2)$, where x_2 is a second state variable that is moderately correlated with x_1 . The thin dashed vector demonstrates the hazard of using local least squares fits when the observation variable y and the state variable x_1 are not functionally related.

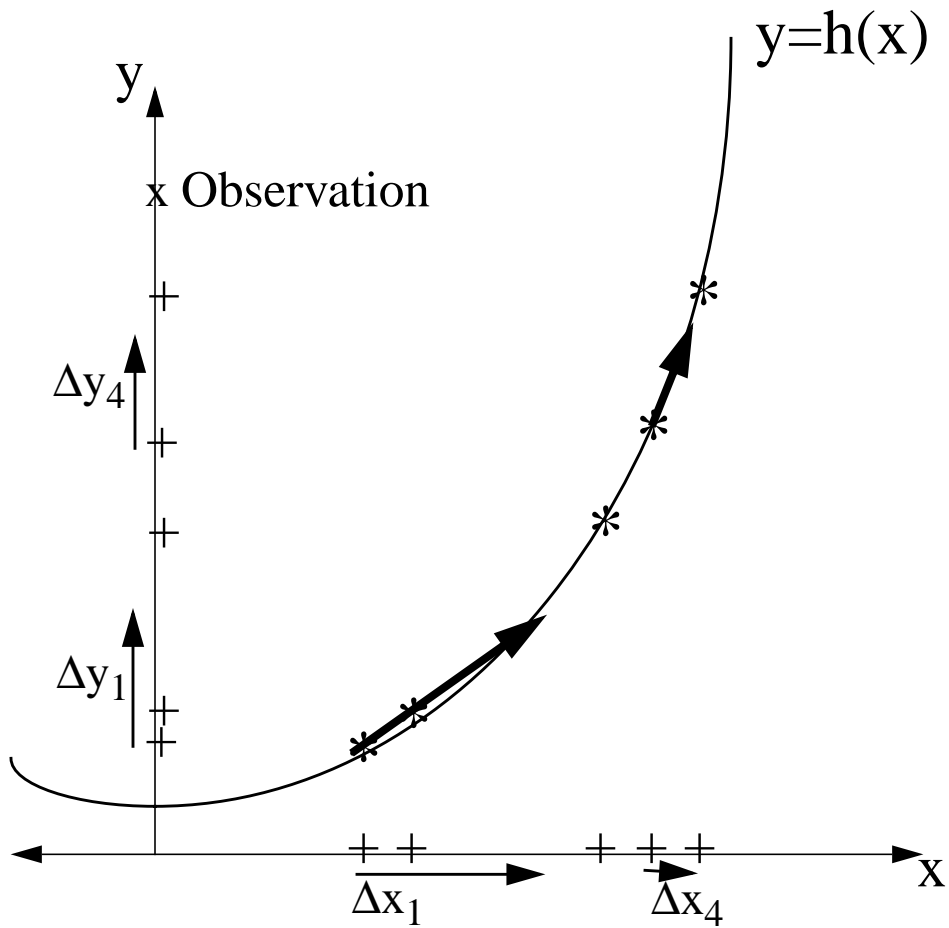
References:

- Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.
- Anderson, J. L., 2001: An ensemble adjustment Kalman Filter for data assimilation. In press with *Mon. Wea. Rev.*
- Andrews, A., 1968: A square root formulation of the Kalman covariance equations. *AIAA J.*, **6**, 1165-1168.
- Bengtsson, T., 2001:
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420-436.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.
- Courtier, P., J. Derber, R. Errico, J.-F. Louis, and T. Vukicevic: 1993: Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus(A)*. **45**, 342-357.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.
- Evensen, G., and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background

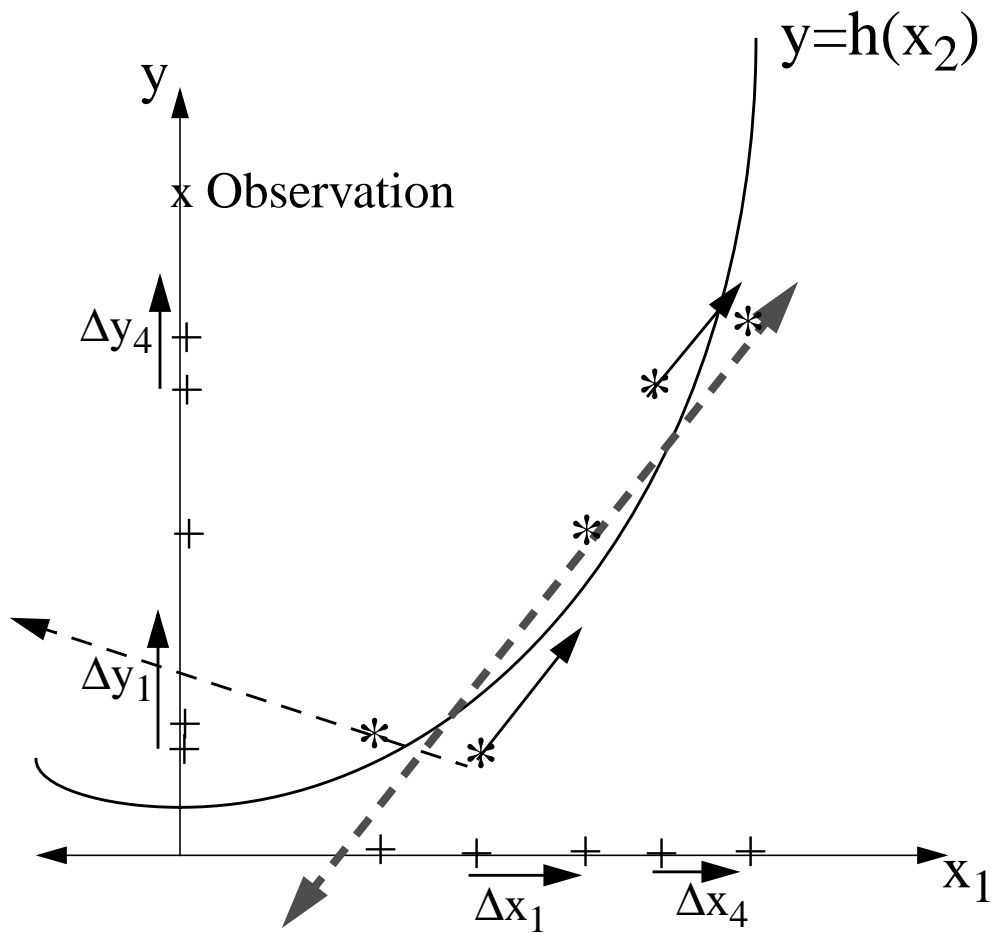
- error covariance estimates in an ensemble Kalman filter. To appear in *Mon. Wea. Rev.*
- Hamill, T. M., and J. S. Whitaker, 2001: Ensemble data assimilation without perturbed observations. In press with *Mon. Wea. Rev.*
- Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.
- Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kalman, R. and R. Bucy, 1961: New results in linear prediction and filtering theory. *Trans. AMSE J. Basic Eng.*, **83D**, 95-108.
- Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971-1981.
- Lermusiaux, P. F. and A. R. Robinson, 1999: Data assimilation via error subspaces statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, **127**, 1385-1407.
- Miller, R. N., M. Ghil and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037-1056.
- Mitchell, H. L. and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.
- Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194-1207.
- Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 175pp.
- Tarantola, A., 1987: *Inverse Problem Theory*. Elsevier Science Publishers B. V., 613 pp.
- Whitaker, J. S. and T. M. Hamill, 2001: Ensemble data assimilation without perturbed observations. In press with *Mon. Wea. Rev.*



1. An idealized representation showing the relation between update increments for a state variable, x , and an observation variable, y , for a five member ensemble represented by asterisks. The projection of the ensemble on the x and y axes is represented by '+' and the observation, y^0 is represented by 'x'. In this case, y is functionally related to x by h . The grey dashed line shows a global least squares fit to the ensemble members. Update increments for ensemble members 1 and 4 for y are shown along with corresponding increments for the ensemble as a whole (thin vectors parallel to least squares fit) and for the x ensemble.



2. As in Figure 1 but showing the application of local least squares fits, in this case using only the nearest neighbor in y , to compute the updates for x given the updates for y . The local updates for the first and fourth ensemble members are shown by the black vectors.



3. As in Figure 1, but now $y = h(x_2)$, where x_2 is a second state variable that is moderately correlated with x_1 . The thin dashed vector demonstrates the hazard of using local least squares fits when the observation variable y and the state variable x_1 are not functionally related.